

Exercise 3

Lecture 5,6: Index Structures & Materialized Views

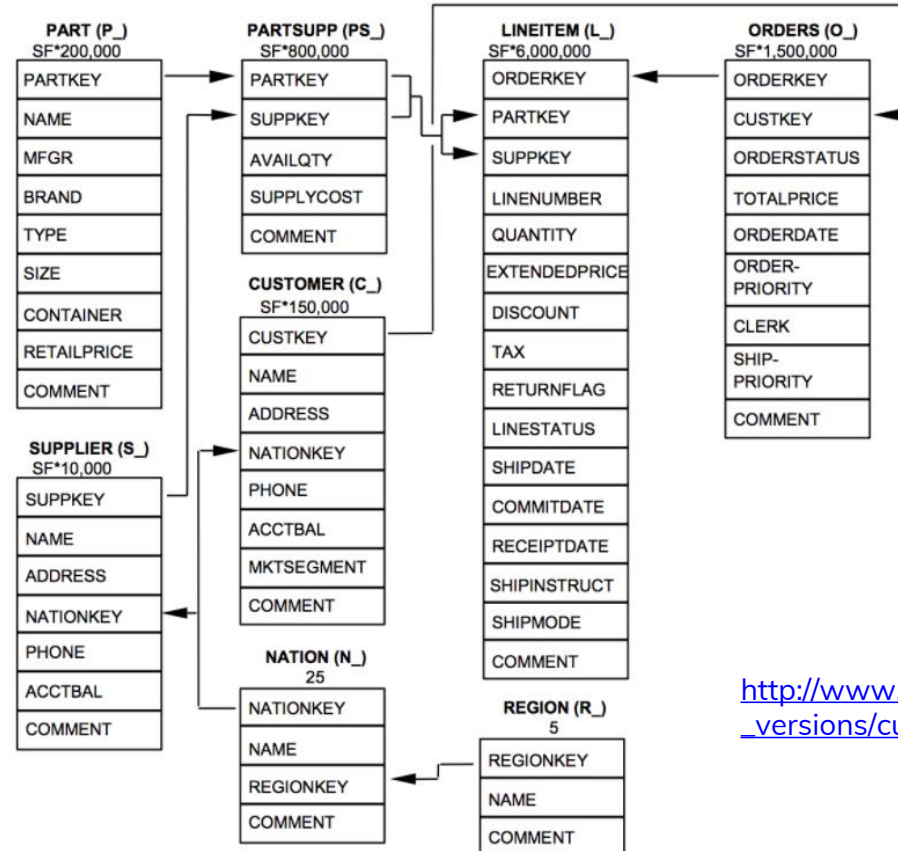
Scalable Data Engineering

1.1 Are the following statements true or false? Give reasons for your answer.

- a) Mat. views can be used to improve execution times of workloads with complex queries.
- b) There is no difference in mat. views containing joins or aggregates.
- c) The creation of compensation queries is an easy, straightforward process.
- d) Deferred maintenance is preferable compared to eager maintenance.
- e) Incremental maintenance is preferable compared to complete maintenance.

The TPC-H Schema

A ROLAP schema



http://www.tpc.org/tpc_documents_current_versions/current_specifications5.asp

1.2 Index creation & use

- a) Make yourself familiar with the explain functionality of pg_admin.
- b) Get all rows of the order relation with the value `clerk#000000322` in the `o_clerk` column and measure the execution time of the SQL statement.
- c) Create a index structure to support filter queries on the `o_clerk` column.
- d) Reissue the SQL statement from b), how did the plan and the execution time change?

1.3 View creation & maintenance

- a) Create a materialized view that calculates the number of customers per region in the TPCB schema.
- b) Insert the tuple (5, 'AUSTRALIA', 'down under') to the region relation and update the view.
- c) Explain what you expect and what is the result.

1.4 View selection & usage

- a) Create the materialized view that supports the following query and measure the improvement. (Hint: try to replace the subquery)

```
SELECT p_name, qty, rank() over(ORDER BY qty
desc)
FROM (
    SELECT p_name, sum(l_quantity) qty
    FROM lineitem, part
    WHERE l_partkey=p_partkey
    GROUP BY p_name
) as r
```

1.4 View selection & usage

- b) Extend the materialized view from 1.4 a) to also supports the following query and measure the improvement. (Hint: focus on common parts of the queries)

```
SELECT p_name, n_name, sum(l_quantity),  
       rank() over(PARTITION BY p_name ORDER BY sum(l_quantity) desc)  
FROM lineitem, supplier, nation, part  
WHERE l_partkey=p_partkey AND l_suppkey=s_suppkey AND  
      s_nationkey=n_nationkey  
GROUP BY p_name, n_name
```